



Development of vocabulary sophistication across genres in English children's writing

Philip Durrant¹ · Mark Brenchley¹

© The Author(s) 2018

Abstract

This paper aims to advance our understanding of how children's use of vocabulary in writing changes as they progress through their school careers. It examines the extent to which a model of lexical sophistication as use of low-frequency, register-appropriate words adequately captures development in vocabulary use across the course of compulsory education in England. We find that the received model needs elaborating in a number of important ways. Specifically: (1) the average frequency of words in the repertoire used by older children is no lower than that of younger children. However, younger children's writing is characterized by extensive repetition of high frequency verbs and adjectives and of low frequency nouns (the latter being a product of a focus on entities which are rarely discussed in adult writing). The role of repetition in this finding implies that lexical sophistication is inseparable from lexical diversity, a construct which is usually treated as distinct. (2) Younger children's writing shows a preference for fiction-like vocabulary over academic-like vocabulary. As they mature, children come to make greater use of academic vocabulary in both their literary and non-literary writing, though this increase is greatest in their non-literary writing. Use of fiction vocabulary remains constant across year groups but decreases sharply in non-literary writing, showing an enhanced sense of register appropriateness. This development of register appropriate word use can be captured by relatively simple frequency-based measures that could readily be employed by teachers and researchers to track writers' development in this aspect of word use.

Keywords Children's writing · Corpus linguistics · Learner corpus · Lexical sophistication · Vocabulary

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11145-018-9932-8>) contains supplementary material, which is available to authorized users.

✉ Philip Durrant
p.l.durrant@exeter.ac.uk

¹ Graduate School of Education, St Luke's Campus, University of Exeter, Heavitree Road, Exeter EX1 2LU, UK

Introduction

This paper aims to advance our understanding of how children's use of vocabulary in writing changes as they progress through their school careers. Specifically, it elaborates on existing models of the features of word use which distinguish the writing of older children from that of younger children. Methodologically, it belongs to a tradition going back to at least the 1930s of studying children's writing development through quantitative analysis of linguistic features. This approach offers a useful complement to qualitative analyses (e.g., Christie & Derewianka, 2008) in that it enables reliable analysis of large numbers of texts, so allowing patterns to emerge which may not be obvious in smaller samples and supporting robust generalizations. The systematicity required of the approach and the reliance on quantitative analysis to identify patterns also enables a distancing of the analyst from the text which can bring out patterns that may not be obvious to the naked eye.

While the majority of studies in this tradition has focused on syntactic development, the last 15 years have seen growing interest in features of vocabulary (e.g., Crossley, Weston, Sullivan, & McNamara, 2011; Malvern, Richards, Chipere, & Duran, 2004; Massey, Elliott, & Johnson, 2005; Olinghouse & Leaird, 2009). Vocabulary development is particularly well-suited to this type of analysis, both because the units of analysis (words) are more numerous than the units of syntax and because they are more easily identified by automated means, so allowing relatively reliable analysis.

The focus on vocabulary has clear practical importance given the emphasis on this as an aspect of writing development in Anglophone school curricula (Australian Curriculum and Assessment Reporting Authority, 2014; Department for Education, 2014; National Governors Association Center for Best Practices, 2010). It is also especially salient given contemporary concerns about the existence of a "vocabulary gap" that is preventing a significant proportion of students from achieving their full potential (Harley, 2018; Quigley, 2018). Such concerns underline the value of explicit descriptions of vocabulary development, both as a means of clarifying what a "vocabulary gap" might actually entail and for ensuring it is effectively targeted.

Quantitative measures of vocabulary development in children's writing

Previous work distinguishes three main types of measure of vocabulary development: measures of lexical density, lexical diversity, and lexical sophistication (Read, 2000). *Density* refers to the proportion of a text which is made up of lexical words (usually defined as verbs, nouns, adjectives and adverbs). This is known to be an important distinguisher of text genres (e.g., Biber, 1988); however, research has shown it to be of little developmental interest (e.g., Berman & Nir, 2010; Uccelli, Dobbs, & Scott, 2013). *Diversity* refers to the repertoire of different words which a writer uses. This is perhaps the most commonly-used

measure of vocabulary development and findings have overwhelmingly supported the conclusion that diversity increases with age (e.g., Berman & Nir, 2010; Crossley et al., 2011; Malvern et al., 2004; Olinghouse & Wilson, 2013; Uccelli et al., 2013).

The literature on *lexical sophistication* is more wide-ranging and offers fewer clear conclusions. Researchers rarely state exactly what they mean by the term, but Read's (2000) definition captures most of what it has been construed as covering. For him, sophistication is the "selection of low-frequency words that are appropriate to the topic and style of the writing, rather than just general, everyday vocabulary" (2000, p. 200).

One operationalization of Read's definition is found in studies which count the proportion of words in a text which are not found on a list of high-frequency vocabulary. Some studies have found this proportion to increase with age (Finn, 1977; Olinghouse & Graham, 2009; Olinghouse & Wilson, 2013; Sun, Zhang, & Scardamalia, 2010), although Malvern et al. (2004) did not find an increase from ages seven to 14, and Lawton (1963) found an increase between 12 and 14 for working class children but not middle-class children. While this method provides an easily-understood measure of sophistication, it is somewhat 'blunt' in that each word receives only a binary score: present or missing from the reference list. A great deal of potentially meaningful variation between more-and-less frequent words on both sides of that divide is thereby lost.

Crossley et al. (2011) take a more comprehensive approach by retrieving from a reference corpus a frequency count for each word in a text and taking the mean of these frequencies to define an overall score for the text as a whole. Using this method, they found no significant difference between ninth and eleventh graders, although college writers did exhibit lower averages than school-level writers. Crossley et al.'s approach has the virtue of finer gradation, it suffers from the fact that word frequencies follow a highly skewed distribution. This is likely to be reflected in strongly skewed frequency profiles within each text, implying that mean frequencies may not provide a good summary of the range of vocabulary a particular writer uses. This may be the reason for the lack of a significant difference between school year groups. Another explanation may be that the study deals only with the top of the range of school years—it is possible that measurable development in vocabulary sophistication has levelled off by ninth-grade.

Fewer studies have focused on the second part of Read's (2000) definition, which refers to appropriateness to the topic and style of writing. Partially relevant here is research which has looked at children's use of Greek- and -Latin-based words (Corson, 1985; Berman & Nir-Sagiv, 2007) and their use of words taken from the *Academic Word List* (Sun et al., 2010), both of which were found to increase with age. These studies indicate an overall movement towards greater use of vocabulary which is typical of an academic or 'learned' style. However, there is no real attempt to establish whether this shift is appropriate to the different kinds of texts that children are writing or to address vocabulary typical of other topics or styles.

In conclusion, while research on lexical diversity and lexical density point to fairly clear conclusions—the former increases as children mature, the latter does not—work on lexical sophistication is more ambiguous. The model which casts

vocabulary sophistication as use of lower frequency, more register-appropriate, words, has strong intuitive appeal but research has not been able to establish that it adequately captures development in children's writing. Results regarding frequency are inconsistent and hampered by overly-simple binary methods which ignore much of the potential variation between texts. Furthermore, the few studies which can be construed as relating to appropriateness have focused on a single style (characterized by academic and Greco-Latin words) and have not attempted to relate use to the different kinds of texts that children write. The present study aims to move work in this area forward by measuring development in vocabulary sophistication across the course of compulsory education in England and exploring how the existing model might be elaborated to provide a more accurate understanding of children's vocabulary development.

Methodology

Corpus

This study is based on a new corpus of children's writing. Texts in the corpus are educationally authentic, in that they were produced as part of children's regular schoolwork, rather than being elicited for research purposes. Schools from across England were contacted by the project team, briefed as to the nature of the project, and invited to participate. All writing was obtained subject to the students' voluntary informed consent, with additional consent obtained from the head teacher, the relevant subject teachers, and the students' legal guardians. The corpus, and related materials, are available for download from the project website <https://gigcorpus.com>.¹

We aimed to collect a set of texts that captures the broad range of writing that students are currently producing during the statutory, or *key*, stages of the English school system. Accordingly, texts were sampled at four points: the ends of Key Stage (KS) 1 (Year 2, when children are 6–7 years old) and KS2 (Year 9, when children are 10–11 years old), encompassing the primary phase of the school system, and the ends of KS3 (Year 9, when children are 13–14 years old) and KS4 (Year 11, when children are 15–16 years old), encompassing the secondary stage. Key stages are intended to constitute coherent educational programmes of learning, with formal assessments undertaken at the end of each. Although the specifics of each stage vary according to both discipline and school, all stages are cued to an overarching 'national curriculum' which specifies the "statutory programmes of study and attainment targets for all subjects" (Department for Education, 2014). Collected between September 2015 and December 2017, the present texts were all produced under the version of this curriculum introduced in 2014 (Department for Education, 2014).

Texts were classified into genres on the basis of their overall purpose. Although various schemas were available for this task (e.g., Nesi & Gardner, 2012; Rose &

¹ As part of our ethical agreement with participating students, registration is required to access this site.

Table 1 Corpus makeup—distribution of texts across year groups, genres and disciplines

	English	Humanities	Science	Genre × year totals
Year 2				
Literary	88			88
Non-literary	219	77		296
Year 6				
Literary	139			139
Non-literary	237	86	156	479
Year 9				
Literary	181			181
Non-literary	243	97	68	408
Year 11				
Literary	57			57
Non-literary	310	44	22	376
Discipline totals	1474	304	246	Total texts: 2024

Martin, 2012), following both a review of the texts and extensive discussion with national curriculum specialists at the university where the research was conducted, we decided to use a bespoke classification. This had three benefits. First, it could be efficiently applied to a large number of texts. Second, it could be consistently applied across the three disciplines within the corpus. Third, it could be consistently applied across the four developmental stages within the corpus. The last point was especially valuable, since it allowed texts to be classified in line with their overarching purpose even if the student was not yet able to demonstrate all generic features required by other schemes.

Our classification is based on a two-way distinction between ‘literary’ and ‘non-literary’ tasks. A ‘literary’ text is one which can be evaluated as successful or unsuccessful without considering any kind of propositional or directive relationship to the world. That is, its contents do not need to be judged as either factually accurate or making a persuasive argument in order for the text to be successful. The primary purpose of a literary text is to be appreciated on its own terms as a piece of stylised writing. Within the present corpus, prototypical examples were *creative fiction* and *literary imitations*.

‘Non-literary’ texts, on the other hand, do need to bear a propositional or directive relationship to an external world in order to be considered successful. Their primary purpose is to (a) accurately depict a particular state-of-affairs, (b) evaluate a particular state-of-affairs, or (c) argue for a particular state-of-affairs to be the case. Prototypical non-literary texts included *autobiographies*, *historical accounts*, *complaint letters*, *literary criticism*, *experimental reports*, and *persuasive speeches*.

Texts were sampled across three disciplines: English, Science, and the Humanities (i.e. History, Geography, and Religious Studies). As can be seen (Tables 1, 2), this approach did not yield a balanced corpus. Partly this reflects the practical difficulty of accessing Science and Humanities departments. However, it also reflects the general distribution of writing across the curriculum, at

Table 2 Corpus makeup—contributors and word counts

	Number of				Words per text		
	Texts	Schools	Writers	Titles	Median	Min	Max
Year 2							
English literary	88	5	62	22	70.5	28	133
English non-literary	219	5	131	18	60	29	133
Humanities non-literary	77	4	75	6	57	29	130
Year 6							
English literary	139	6	111	23	293	100	415
English non-literary	237	7	131	20	180	99	412
Humanities non-literary	86	4	70	11	197.5	99	406
Science non-literary	156	4	84	10	207.5	99	396
Year 9							
English literary	181	4	181	6	324	91	522
English non-literary	243	7	134	28	274	92	517
Humanities non-literary	97	3	39	16	236	94	507
Science non-literary	68	3	39	22	187	94	517
Year 11							
English literary	57	4	44	7	355.7	133	630
English non-literary	310	6	96	58	313.5	123	646
Humanities non-literary	44	1	19	8	229.5	124	556
Science non-literary	22	2	22	5	346.5	130	635
Overall	2024	24	828	258	217	28	646

least in terms of ‘continuous prose’, which was the intended focus of the corpus. Thus, the predominance of English texts plausibly reflects the marked emphasis of this discipline on the production of continuous prose; the lack of Year 2 Science texts plausibly indicates a tendency of continuous prose to be a later-developing feature of school Science; and the lack of ‘literary’ Humanities and Science texts reflects these disciplines’ emphasis on dealing with the external world (see below for definitions and discussion of our genre categories).

Once catalogued, texts were typed up and checked by a team of transcribers. Text was removed where it might directly identify either the student or another individual connected with them/the school. Where possible, such material was replaced with an anonymisation marker; where such replacements were not possible, the sentence in which the material occurred was excised in full. In the version of the corpus used in this study, spelling and capitalization were regularized to the conventions of Standard British English. End-of-sentence punctuation was also regularized.

The full corpus comprises 2901 texts. For the present study, however, certain texts were excluded. Specifically:

- Texts that did not constitute continuous prose (e.g., labelled diagrams, sentence exercises, poetry)

- Texts that had a high proportion of illegible words; specifically, any texts with more than 10% illegible words
- Texts that were unusually short or long, in relation to other writing in their year group; specifically, any texts more than one standard deviation above or below the mean for their year group. There were two reasons for this: (a) some year 2 texts were too short for a meaningful analysis to be conducted; (b) previous research has shown text length to be a strong predictor of quality (Bartlett, 1984; Crossley, Roscoe, & McNamara, 2014; Koutsoftas & Gray, 2012), implying that unusually long or short texts may include language which is untypical of their age group.

The makeup of the resulting corpus is shown in Tables 1 and 2. It comprises 2024 texts representing 258 distinct titles, written by 828 children from 24 different schools. Text length tends to increase across year groups, and literary texts tend to be longer than non-literary texts. Texts are reasonably evenly-split across genders, with 52.9% written by females, 42.9 written by males and the remainder unknown. 20.2% of texts were written by pupils eligible for students deemed eligible for special funding due to their disadvantaged socio-economic status. This figure is slightly above the that for the population—14.1% in state-funded primary schools and 12.9% in state-funded secondary schools (Department for Education, 2017). 12.9% of texts were written by students classified as speaking English as an Additional Language (EAL), slightly below the proportions in the population—20.6% in state-funded primary schools and 16.2% in state-funded secondary schools (Department for Education, 2017). The official definition of EAL used in schools is that students have been “exposed to a language at home that is known or believed to be other than English” (Department for Education, 2017, p. 10). The Department of Education emphasises that EAL status is in no way “a measure of English language proficiency or a good proxy for recent immigration” (Department for Education, 2017, p. 10) and our own experience in working with these texts confirms that it is not a meaningful linguistic category.

Like many corpora, the texts that form the data points in our analyses are not independent: for example, multiple texts are written by individual writers and multiple writers are sampled from individual schools. As Gries (2015) has argued, data of this sort violate the assumption of independence on which standard statistical methods are based. Separate texts written by a single writer or to a single title, or produced within a single school or subject area, are clearly more closely related to each other than they are to those produced by another writer, to another title, or in another school or subject area. Moreover, it is plausible that each of these grouping variables (i.e. writers, titles, disciplines, schools) has the potential to exert its own influence on vocabulary use. To address these issues, our analyses follow recent corpus linguistic practice (Gries, 2015) in making use of mixed-effects models. Such models have two virtues (Tabachnick & Fidell, 2014; Zuur, Ieno, Saveliev, & Smith, 2009). First, they overcome the problem of non-independence, specifically factoring the non-independence of our data into the models. Second, they enable us to better determine the actual significance of any predictors, effectively “re-calculating” the final regression line so as to account for the wider impact of our grouping variables.

Reference data

The following analyses draw on the detailed frequency listing of 100,000 words created by Davies (2018). The version of Davies's list used in this study was accessed in November 2012 and includes frequencies of words in several different corpora and in specific registers within those corpora. For the present study, we used frequencies from the Corpus of Contemporary American (COCA) (Davies, 2008–). Although our study focuses on children in England, this was considered a more relevant and reliable reference point than the British National Corpus (BNC) both because it is more contemporary (collection of texts for the BNC ceased in the early 1990s) and because it is substantially larger (450 million words, in comparison to 100 million words) and covers a greater number of word types (10% of words from the 100 K COCA list are not found in the parallel BNC-based list). We assume that, in spite of minor differences that could be cited for a few individual words, frequencies in American and English contexts are likely to be highly correlated and hence that the American origin of the frequency lists will have a negligible influence on our results. Indeed, a simple correlation analysis of COCA- and BNC-based lists (excluding items not found in the BNC) shows a correlation of $r_s = .82$. Proper names, numbers and units of measurement are not included in the COCA list, so will not form a part of the analyses which follow.

In choosing COCA as a reference, we are deliberately defining vocabulary sophistication in terms of texts' relationship to adult discourse. This approach rests on the assumption that sophistication should be gauged with reference to the sorts of discourse towards which children's education aims (what we might call a *teleological* approach to defining sophistication). The obvious alternative would be to use a corpus of the sorts of discourse to which children at particular ages are likely to have been exposed (e.g., age-appropriate school textbooks or children's fiction). This would certainly be a worthwhile exercise, giving valuable information about the relationship between the language which children use and the language to which they are exposed. However, this backward-facing reference point (what we might call a *causal* approach) would, we believe, be less useful as a way of defining sophistication. This is both because sophistication, in our view, should focus on the goals towards which children are aiming, rather than on where they have come from, and because the multiple reference corpora that would be needed to study children across different age groups would not provide a consistent point of reference against which development could be understood. It should be born in mind throughout this paper that the terms 'low/high-frequency' mean low/high-frequency in comparison with adult norms. This follows the practice of the previous research on sophistication described above.

Processing the study corpus

The study corpus was first tagged for part of speech using CLAWS (Garside & Smith, 1997). Because the COCA frequency lists employ a slightly simplified

version of CLAWS's C7 tagset, tags were post-edited using a search-and-replace script to match those used in the COCA lists.² To enable comparison with the COCA frequency lists, British English spellings were converted to US spellings using the comprehensive list available at <http://www.tysto.com/uk-us-spelling-list.html>.³ Frequencies for each word in each text of the study corpus were then retrieved from the COCA list. Specifically, for each word, we recorded total occurrences per million words and occurrences per million words in each of its five register sub-corpora (*spoken, fiction, newspaper, magazine, academic*). Because use of function words is likely to reflect differences in syntactic structures, rather than differences in vocabulary *per se*, our analyses are based on counts only of adjectives, adverbs, nouns, and lexical verbs.

A central issue in any study of vocabulary frequency relates to how individual words should be defined. The simplest approach is to count any identically-spelt items as examples of the same word. While this is computationally easy to implement, it has the double disadvantage of conflating some things that we might wish to distinguish (e.g., the high-frequency noun *address* and the much lower-frequency verb *address* would be recorded as the same word) and distinguishing things that we might wish to treat together (e.g., the base verb *argue* would be treated as a distinct word from its inflected forms *argues*, *argued* and *arguing*). Three alternative are readily available (see Gardner, 2008 for discussion):

1. Non-lemmatized approach: Treat word form-part of speech combinations as distinct words. For example, *address* (noun), *addresses* (noun), *address* (verb), *addresses* (verb) would each be counted as distinct words. This is a relatively fine-grained approach, achieving maximum distinctions between different words.
2. Lemmatized approach: Combine inflected forms of words within a single part of speech. Thus, the plural and singular forms of *address* as a noun would be treated as one word and the various inflections of *address* as a verb would be treated as another.
3. Word-family approach: Treat both inflectional and derivational variations as a single item. On this approach, all forms of *address* and both verb and noun would be treated as the same item, along with the derived noun *addressee*.

We believe that option 3 is too broad-brushed to produce a meaningful analysis, often conflating words which may not have clear links between them for writers [e.g., Coxhead's *Academic Word List* (2000), which took this approach, counts as a single item such diverse forms as *constitute*, *constituency* and *unconstitutional*]. However, there are no obvious *a priori* reasons for believing that either 1 or 2 provides the most relevant information. In the analyses which follow, data will be shown for both lemmatized and non-lemmatized frequencies. As will be seen, the two sets of data provide very similar descriptive findings. To avoid multiplying

² This, and all further text edits and statistical analyses, was implemented using *R* (R Core Team, 2013. *R*: A language and environment for statistical computing).

³ This list was slightly amended to include the frequent child-writing item *mum* (corrected to US *mom*).

inferential analyses, we have therefore performed inferential tests only for the non-lemmatized data (i.e. option 1).

A second issue relates to whether analysis should count word *tokens* (i.e. all words, regardless of whether they have been used before in the text) or word *types* (i.e. distinct words, ignoring repeated uses of the same word). As noted above, previous research has shown that younger children tend to repeat words more than older children, raising the possibility that analyses based on type and token counts will provide usefully different perspectives. Accordingly, both token- and type-based counts will be presented in the following analyses.

Inferential methods

As mentioned above, our texts require the use of mixed-effects models. Accordingly, for each analysis, we adopted the three-stage stepwise procedure detailed in Gries (2015) and outlined below.⁴

Stage One involved identifying the maximal fixed effects structure and the maximal random effects structure of interest. For all analyses reported below, the maximal fixed effects structure comprised the main effects of year group and genre plus their interaction. Conversely, the maximal random effects structure comprised two crossed sets of nested effects, yielding four random effects overall: schools; disciplines; writers as nested within schools; titles as nested within disciplines. The two nested structures are crossed because individual titles were written by multiple writers, whilst individual writers wrote on multiple titles. Titles also cut across schools as students from multiple schools wrote on common titles, reflecting the influence of a national curriculum with shared public examinations.

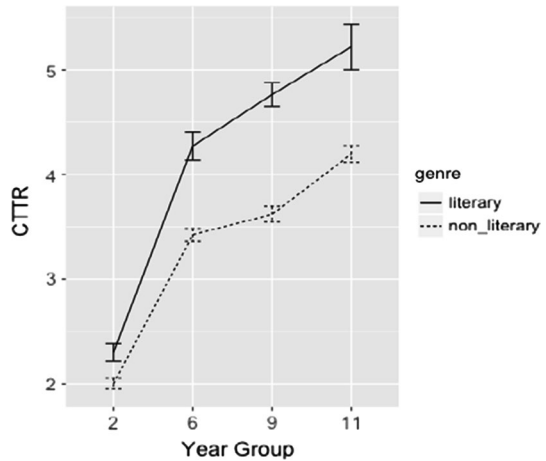
For Stage Two, we combined the maximal fixed effects structure with the maximal random effects structure. We then determined the optimal random effects structure relative to this combination by (a) removing each random effect in turn, and (b) comparing the overall quality of the model when the effect is present versus when it is absent. In each case, particular random effects were retained only if their removal made the model quality significantly worse; otherwise, the effect was eliminated from the final Stage Two model altogether.

For Stage Three, we determined the optimal fixed-effects structure relative to the optimal random effects structure identified in Stage Two. This involved sequentially removing any fixed effects which were neither significant in themselves nor participated in any higher order interactions. As with the Stage Two procedure, a particular fixed effect was retained only if removing it made the model quality significantly worse; otherwise, the effect was eliminated in order to derive the final models reported below.

In both stages, model quality was determined with reference to the Akaike Information Criterion (AIC) score for each model iteration. This is an estimate of model

⁴ All models were implemented using R version 3.2 and the *lmerTEST* R package (Kuznetsova, Brockhoff, & Christensen, 2017), with goodness of fit statistics calculated using the MuMIn package (Barton, 2018).

Fig. 1 Corrected type-token ratio



quality well-suited to exploratory analysis, identifying the model that best predicts the values of future samples (Aho, Derryberry, & Peterson, 2014).

Finally, as an extension of standard linear regression, mixed-effects models need to meet certain assumptions to be accurate and generalizable (Tabachnick & Fidell, 2014; Zuur et al., 2009). These were checked as follows: histograms of residuals were checked to identify significant outliers; residuals versus observed values were checked to confirm the linearity of the data; Q–Q plots were checked to confirm the normal distribution of residuals; plots of standardized residuals versus fitted values were checked to confirm homoscedacity of residuals. All analyses met the necessary assumptions.

Analysis

Preliminary analysis: vocabulary diversity across year groups

One of the strongest findings of previous research has been that children use a wider range of vocabulary with age. Though it is not a focus of the current study, measuring vocabulary diversity within our corpus will be important for interpreting the main analysis. To quantify this, we used the corrected type-token ratio (CTTR), a variation on the traditional type-token ratio which allows reliable comparisons across texts of different lengths (Carroll, 1964). CTTR is calculated as (non-lemmatized) types (distinct words) divided by the square-root of twice the token (total words) count and higher scores show greater diversity. CTTR across year groups and text genres in the study corpus is shown in Fig. 1. As expected, literary texts were more diverse than non-literary and diversity increased with age, trends confirmed

Table 3 Mixed-effects model for CTTR

	Value	SE	<i>df</i>	<i>t</i> -value	<i>p</i> value
Intercept	1.785	.252	76.47	7.081	<.0001
Year group	.318	.026	167.70	12.264	<.0001
Genre	−.700	.087	243.59	−8.041	<.0001
Random effects			Variance		SD
Writer within school			.126		.355
Title within discipline			.199		.446
School			.350		.591
Residual			.191		.437
Goodness of fit					
R^2 marginal			.55		
R^2 conditional			.90		

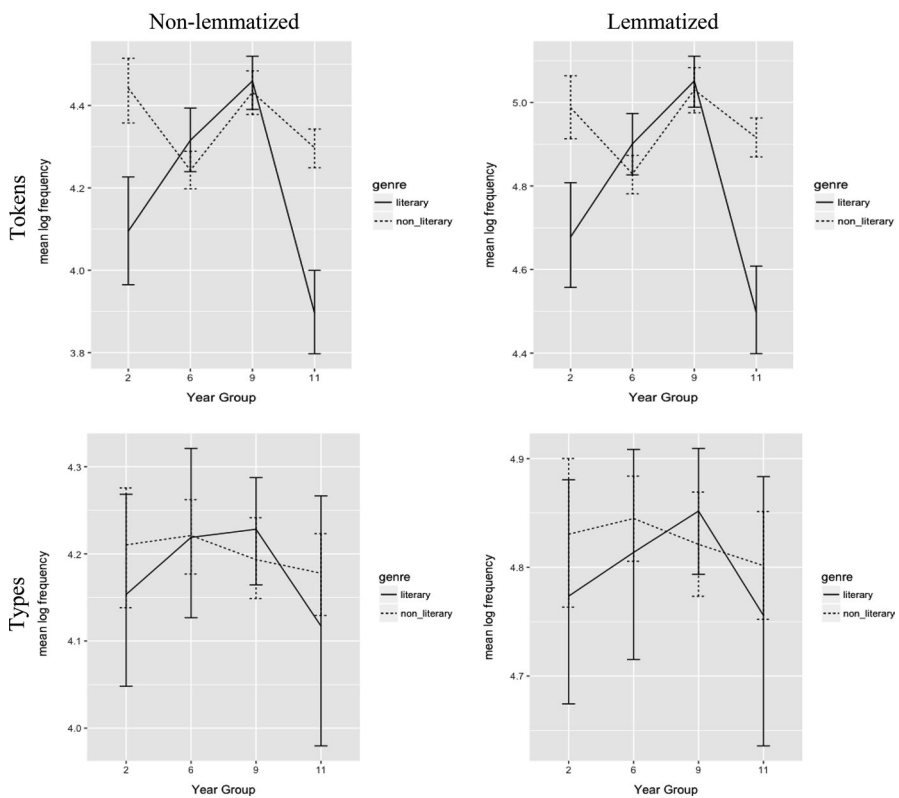
**Fig. 2** Mean log frequencies for all parts of speech

Table 4 Mixed-effects model for non-lemmatized tokens

	Value	SE	<i>df</i>	<i>t</i> -value	<i>p</i> value
Intercept	4.021	.220	2.36	18.273	.0013
Genre	.235	.066	291.27	3.539	.0005
Random effects	Variance			SD	
Writer within school			.019		.139
Title within discipline			.114		.337
School			.023		.151
Discipline			.127		.356
Residual			.116		.340
Goodness of fit					
R^2 marginal			.02		
R^2 conditional			.72		

Table 5 Mixed-effects model for non-lemmatized types

	Value	SE	<i>df</i>	<i>t</i> -value	<i>p</i> value
Intercept	4.199	.046	33.90	90.460	<.0001
Genre	−.045	.030	1759.90	−1.494	.135
Random effects	Variance			SD	
Writer within school			.028		.166
School			.030		.174
Residual			.205		.453
Goodness of fit					
R^2 marginal			.00		
R^2 conditional			.22		

by the mixed-effects model shown in Table 3. Example texts which are close to the mean CTTR figure for their year group \times genre combination are provided in the supplementary materials⁵ (Part A).

Frequency profiles

The procedure outlined in the methodology section provides a frequency value for each word in each text of the study corpus. The analytical challenge is to provide an informative and intuitively comprehensible summary of this rich information. As noted above, skewed frequencies of words within each text make the

⁵ Available from https://phildurrantnet.files.wordpress.com/2018/12/readwrit_supplementary_materials.pdf.

mean a poor summary. We therefore used log frequencies, which provide a more normal distribution within each text. Figure 2 shows the mean of mean log frequencies across year groups and genres for all lexical words. Tables 4 and 5 show the best-fitting mixed-effects models for the non-lemmatized data. No clear patterns are visible across year groups for either analysis. In the analysis of types, mean frequencies were lower in literary than in non-literary writing.

The fact that mean word frequency does not decrease across year groups is surprising, and leaves us with a choice between three conclusions:

1. Vocabulary sophistication does not increase as children progress through schooling.
2. Vocabulary sophistication is not related to frequency.
3. Our current measure of frequency is not sufficiently sensitive to capture decreases in frequency.

Of these, option three appears the most plausible. We therefore developed a more fine-grained picture of vocabulary looking separately at each part of speech. Figure 3a–d and Tables 6, 7, 8, 9, 10, 11, 12 and 13 show data and best-fitting models separately for adjectives, adverbs, nouns and verbs. As lemmatized and non-lemmatized versions appear to be parallel, inferential tests were run only for non-lemmatized versions. Together with the two models shown in Tables 4 and 5, these bring the total number of analyses in this part of the paper to 10. A conservative threshold of $.05/10 = .005$ is therefore adopted for statistical significance. Representative texts which are close to the mean figure for their year group \times genre combination for each part of speech are provided in the supplementary materials.²

Four points stand out from these data. First, in the token-based analyses, all parts of speech show significant differences across year groups. Second, the differences across year groups seen for nouns is in the opposite direction to that of the other parts of speech. That is, while the mean frequency of other parts of speech decreases as age increases, the mean frequency of nouns increases. Presumably, this divergence between nouns and the other parts of speech is the reason why no effect for age could be seen in the analysis of all parts of speech. Third, in the token-based analyses, mean frequency of adjectives, adverbs and verbs is significantly lower in literary than in non-literary texts. Again, nouns buck the trend, not showing any significant difference between genres. Finally, analyses based on types do not show significant effects for either year group or genre in any part of speech.

The higher percentage of low-frequency verbs and adjectives in literary versus non-literary texts is in line with our expectations, as is the increased proportion of such words as children get older. However, the increase in noun frequency as children progress through school and the lack of significant effects in the analyses by types are unexpected and require further discussion. Table 14 shows excerpts from year 2 and year 11 literary and non-literary texts. Each excerpt comes from a text that is close to the mean for noun use in its category. Nouns with frequencies of below 10/million words in COCA are underlined.

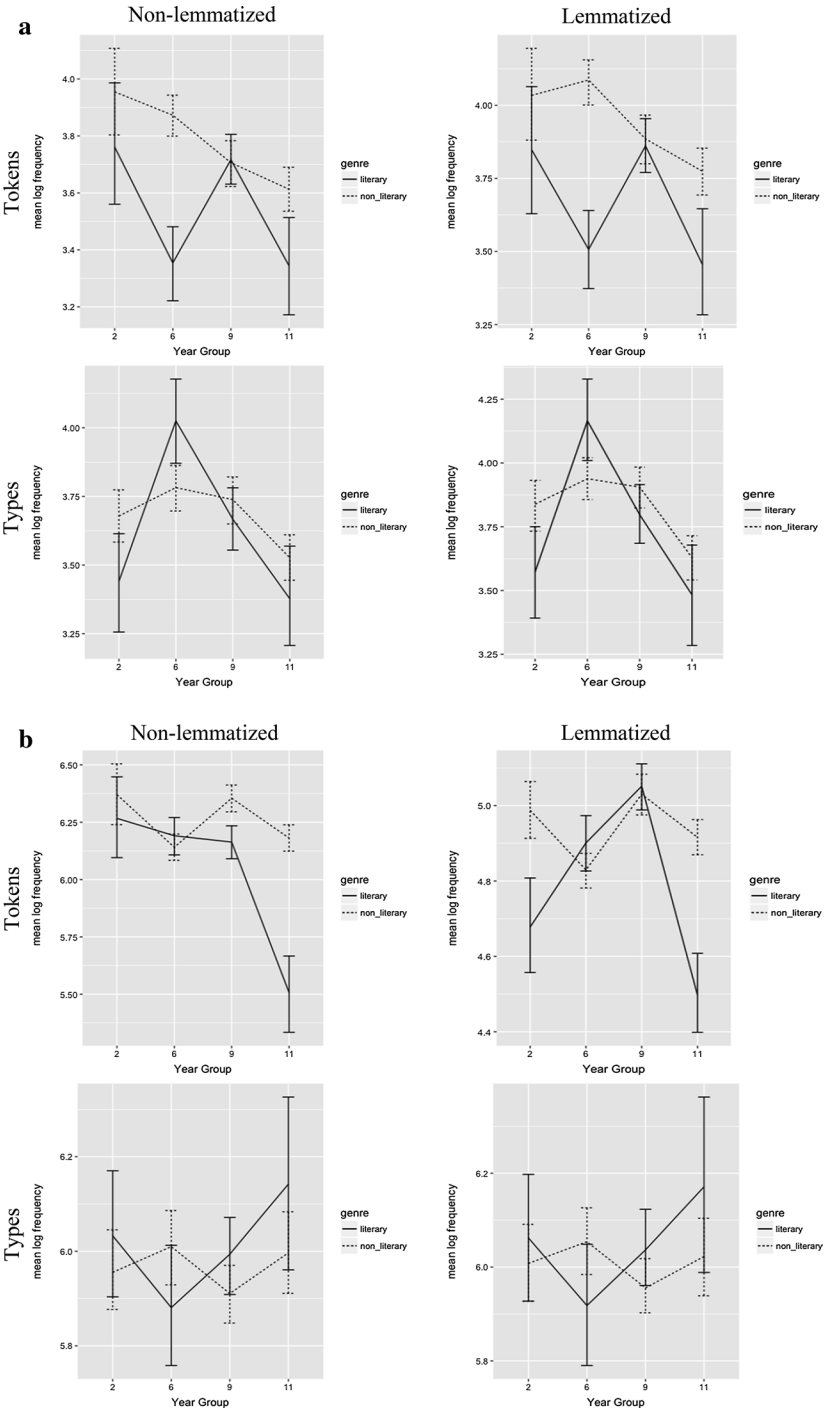


Fig. 3 Mean log frequencies for **a** adjectives, **b** adverbs, **c** nouns, **d** verbs

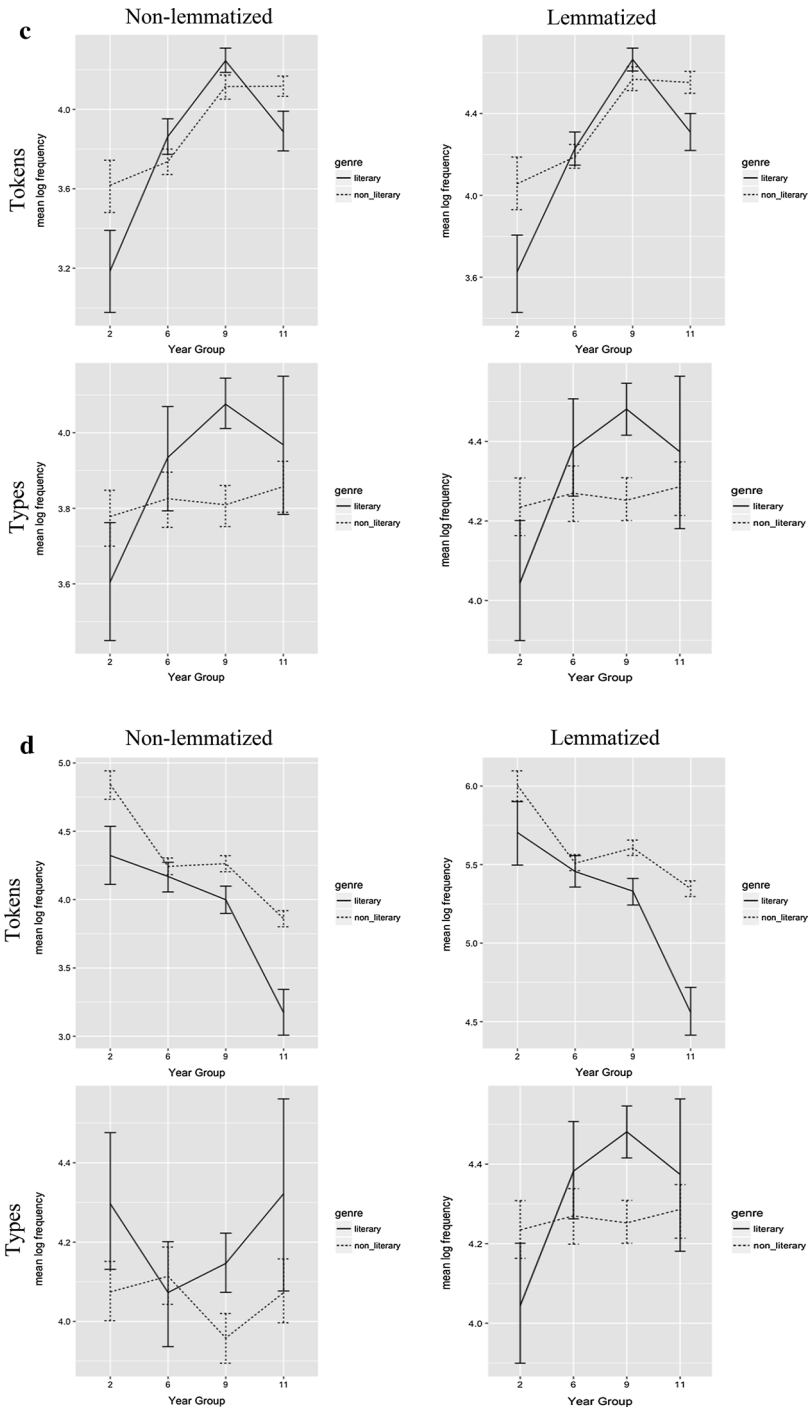


Fig. 3 (continued)

Table 6 Mixed-effects model for non-lemmatized adjective tokens

	Value	SE	df	t-value	p value
Intercept	3.852	.198	4,510	19.418	<.0001
Year group	–.058	.014	220.71	–4.233	<.0001
Genre	.352	.109	219.08	3.231	.0016
Random effects	Variance			SD	
Writer within school			.038		.194
Title within discipline			.251		.501
Discipline			.067		.258
Residual			.528		.726
Goodness of fit					
R^2 marginal			.05		
R^2 conditional			.43		

Table 7 Mixed-effects model for non-lemmatized adjective types

	Value	SE	df	t-value	p value
Intercept	3.589	.152	69.50	23.617	<.0001
Year group	.005	.020	126.46	.277	.783
Genre	.042	.139	189.37	.304	.762
Year group \times genre	–.003	.020	218.39	–.148	.882
Random effects	Variance			SD	
Writer within school			.043		.207
Title within discipline			.021		.146
School			.033		.180
Residual			.666		.816
Goodness of fit					
R^2 marginal			.00		
R^2 conditional			.13		

These excerpts (and those presented in the supplementary materials²) show a pre-occupation in the younger children's writing with entities that are relatively unusual from the perspective of the adult discourse represented in our reference corpus. The fiction, newspaper, magazine and academic texts which make up the COCA generally have much less interest in *fairies*, *playtime* and *dinosaurs* than do the young writers in the lower years of our study corpus. However, this tendency towards distinctively 'child-like' topics cannot be fully explain our findings. As was noted above, the overall repertoire of nouns used (as shown in the analysis by types) does not vary significantly across year groups, so older children are just as likely to use infrequent nouns as younger children. The key difference between year groups lies, rather, in the prominent role which infrequent nouns play due to their extensive

Table 8 Mixed-effects model for non-lemmatized adverb tokens

	Value	SE	<i>df</i>	<i>t</i> -value	<i>p</i> value
Intercept	6.472	.160	62.47	40.41	< .0001
Year group	− .069	.022	126.90	− 3.20	.0017
Genre	− .030	.160	219.72	− .19	.8512
Year group × genre	.043	.022	236.15	1.95	.0527
Random effects	Variance			SD	
Writer within school			.096		.310
Title within discipline			.030		.174
Residual			.422		.650
Goodness of fit					
R^2 marginal			.05		
R^2 conditional			.27		

Table 9 Mixed-effects model for non-lemmatized adverb types

	Value	SE	<i>df</i>	<i>t</i> -value	<i>p</i> value
Intercept	5.907	.113	149.00	52.126	< .0001
Year group	.007	.015	238.50	.448	.655
Genre	.093	.102	1894.80	.916	.360
Year group × genre	− .011	.014	1114.80	− .767	.443
Random effects	Variance			SD	
Writer within school			.035		.186
School			.016		.128
Residual			.491		.701
Goodness of fit					
R^2 marginal			.00		
R^2 conditional			.10		

repetition. This is illustrated well in Table 14. While the low-frequency nouns in the Year 11 texts appear only once each, four of the six low-frequency nouns in the Year 2 literary text are forms of the lemma *fairy* and two of the four low-frequency nouns in the non-literary texts are variants of *turtle*. Changes in noun use, it seems, are not best captured in the repertoire of words which are used, rather there is a tendency for younger children to heavily recycle lower-frequency items.

A parallel description can be given for the other parts of speech, where again the overall repertoire of adjectives, adverbs and verb does not change across year groups, but repetition of high-frequency items in younger children's writing gives way to more diverse use amongst older children (see Part B of the supplementary materials² for illustrations). We have already seen from our preliminary analysis that vocabulary use in our corpus becomes more diverse across year groups—that

Table 10 Mixed-effects model for non-lemmatized noun tokens

	Value	SE	<i>df</i>	<i>t</i> -value	<i>p</i> value
Intercept	3.444	.281	2.92	12.259	.0013
Year group	.050	.015	20.05	3.358	.0031
Random effects	Variance			SD	
Writer within school			.014		.120
Title within discipline			.244		.494
School			.020		.140
Discipline			.185		.430
Residual			.182		.427
Goodness of fit					
R^2 marginal			.04		
R^2 conditional			.73		

Table 11 Mixed-effects model for non-lemmatized noun types

	Value	SE	<i>df</i>	<i>t</i> -value	<i>p</i> value
Intercept	3.656	.121	169.50	30.101	<.0001
Year group	.0027	.015	512.70	1.838	.0667
Genre	.0157	.089	1971.70	.175	.8607
Year group \times genre	-.0105	.013	1775.30	-.822	.4114
Random effects	Variance			SD	
Writer within school			.042		.206
School			.049		.222
Residual			.363		.603
Goodness of fit					
R^2 marginal			.01		
R^2 conditional			.21		

is, that younger children's vocabulary is more repetitive than older children's writing. It is now clear that this repetition interacts with frequency effects to produce the significant results seen above.

Taken together, these findings suggest two refinements to the model of vocabulary sophistication as selection of low-frequency words. First, different parts of speech show radically different developmental profiles so at least this minimal level of syntactic information needs to be incorporated into our vocabulary models. Second, younger children's writing is distinguished from that of older children, not by a repertoire of words that is more-or-less frequent, but rather by greater repetition of low-frequency nouns and high-frequency adjectives, adverbs and verbs. On this view, vocabulary diversity and vocabulary sophistication are

Table 12 Mixed-effects model for non-lemmatized verb tokens

	Value	SE	<i>df</i>	<i>t</i> -value	<i>p</i> value
Intercept	4.538	.172	10.35	26.39	<.0001
Year group	–.102	.015	32.29	–6.720	<.0001
Genre	.401	.085	230.58	4.741	<.0001
Random effects	Variance			SD	
Writer within school			.029		.173
Title within discipline			.152		.390
School			.027		.164
Discipline			.027		.164
Residual			.312		.558
Goodness of fit					
R^2 marginal			.18		
R^2 conditional			.54		

Table 13 Mixed-effects model for non-lemmatized verb types

	Value	SE	<i>df</i>	<i>t</i> -value	<i>p</i> value
Intercept	4.208	.147	135.30	28.641	<.0001
Year group	–.018	.017	573.30	–1.046	.296
Genre	–.057	.010	1930.00	–.573	.567
Year group \times genre	–.026	.014	1792.00	–.179	.858
Random effects	Variance			SD	
Writer within school			.066		.258
School			.096		.310
Residual			.442		.665
Goodness of fit					
R^2 marginal			.00		
R^2 conditional			.27		

not—as previous research has construed them—separate constructs, but rather interact to distinguish writing at different levels.

Appropriateness

It will be recalled from “[Methodology](#)” section that the second part of Read’s definition of lexical sophistication refers to “words that are appropriate to the topic and style of the writing” (Read, 2000: 200). We operationalize this through the register-specific frequency counts provided in the COCA frequency lists. Separate frequency counts (normalized to occurrences per million words) are provided for five sub-corpora within COCA: *spoken*, *academic*, *fiction*, *magazine*, and *newspapers*.

Table 14 Excerpts from texts with close to mean scores for noun frequency

	Literary	Non-literary
Year 2	<p>One day on a stormy, wet, cold, morning Rosie saw her first red <u>fairy</u>. She looked up and saw a magical <u>fairy</u>. She let the red <u>fairy</u> in the house. After <u>playtime</u> the more Rosie looked the more book <u>fairies</u> she saw in the sky but no one else noticed them. Rosie went into the Institutionname class room and in the <u>trays</u> she found a map to <u>fairyland</u></p> <p>(6_623 h, mean noun log frequency = 3.2)</p>	<p>I am writing to you about the sea <u>turtle</u> because they are not safe. People like the fishermen throw <u>nets</u> in the sea and if <u>turtles</u> get stuck in them they will ILLEGIBLE_TEXT die. Also that can happen to other animals in the sea and at the beach. People throw pollution in the sea but they are hunted for their shells and they're killed just for their shells. They've been alive since the <u>dinosaurs</u> were alive and hunters can kill them...</p> <p>(15_803c, mean noun log frequency = 3.6)</p>
Year 11	<p>The monotonous, shrill <u>screech</u> of the alarm clock brought me to my senses, as I wearily stumbled out of my bed and into the bathroom. A <u>shroud</u> of darkness lingered outside, accompanied by the persistent <u>patter</u> of rain. As I looked through the window, dark clouds slowly circled around, menacing and patient. I caught the 7:21 train on platform 3, like I do every day and the familiar scent was oddly welcoming</p> <p>(22_1066f, mean noun log frequency = 3.9)</p>	<p>At the start of the play when the inspector is hinting at the <u>accusation</u> that the family is responsible for the death of Eva Green. Sheila immediately questions the inspector. Saying "you talk as of we are responsible". This comes across childish and it's as if it's almost impossible that they are related to the incident. Before that when Gerald is proposing to her, a big moment in anyone's life, she doesn't seem to take it very seriously</p> <p>(23_1118c, mean noun log frequency = 4.1)</p>

Table 15 Sample of genre proportions from the transformed COCA frequency list

Word form	POS	Academic	Fiction	Magazine	News	Spoken
The	Article	.24	.20	.20	.20	.17
And	Coordinating conjunction	.23	.19	.20	.18	.20
Of	Preposition	.30	.15	.20	.18	.17
Shuddered	Past tense lexical verb	.01	.90	.06	.02	.01
Tunelessly	General adverb	.03	.90	.08	.00	.00
Metacognitive	General adjective	1.00	.00	.00	.00	.00
Reflexivity	Singular noun	.98	.00	.02	.00	.00

We use these counts to determine how characteristic individual words are of each register. Specifically, for each word in the COCA list, the five register-based frequency counts are summed to create a total figure representing corpus frequency (per five-million words). The frequency for each register is divided by this total to create a figure representing the proportion of uses of a word which are found in that register. Thus, each register frequency is transformed into a number between zero and one, with the sum of the five numbers totalling to one. If a word is evenly distributed across the five registers, each will have a value of .2. If a word is exclusive

to a single register, that register will have a proportion of one and all other registers will be zero. Table 15 exemplifies these figures for a small sample of words.

In the analyses that follow, we assume that *appropriate* vocabulary involves use of words that score highly within the register of the text being written. We also assume that the *literary* texts in our corpus are closest in target style to the fiction register, while the *non-literary* texts are closest in target style to the academic register. We therefore expect more sophisticated literary texts to use words which score highly on the COCA fiction scale and more sophisticated non-literary texts to use words which score highly on the COCA academic scale. It should be noted that this notion of appropriateness does not address the question of whether a word is used accurately or not (i.e. whether it captures the intended meaning). Rather, the focus is on whether words match the target register.

To quantify this, each lexical word in each text in our corpus was assigned scores from the fiction and academic COCA scales and a mean score on each scale calculated for each text, representing its overall orientation towards the two registers. The mean scores for each year group \times genre are shown in Fig. 4a, b. Like the analysis of frequency, no obvious differences exist between the analyses for lemmatized and non-lemmatized analyses. Unlike the frequency analysis, there are also no obvious differences between token and type analyses. Because of the strong parallels between these four sets of data, inferential statistics were employed only once—for the non-lemmatized type-based analysis. The mixed-effects models related to these are shown in Tables 16 and 17. Because there are two analyses, a conservative alpha of $.05/2 = .025$ is adopted.

Two key developmental patterns are evident. Firstly, vocabulary becomes more academic in style as children progress through school. Secondly, there are significant interactions between year group and genre for both vocabulary types. These reflect the facts that (a) the increase in academic style is more marked in non-literary than in literary texts, and (b) use of fiction words remains relatively constant in literary texts but decreases in non-literary texts. Both patterns suggest an overall shift towards more register-appropriate word use.

Additionally, it is worth emphasising that the goodness of fit achieved by these models exceeds those achieved by the frequency models described in the previous section. Marginal R^2 s (i.e. the percentage of variance accounted for by the fixed effects of year and genre) are .31 and .41. In comparison, figures for the simple frequency models were .04 (nouns), .05 (adjectives and adverbs) and .18 (verbs). The register-based measures, which have traditionally been neglected in studies of vocabulary sophistication, therefore appear to be far more reliable indices of development than purely frequency-based measures.

Discussion and conclusions

This paper has looked at two aspects of lexical sophistication: use of low-frequency words and use of words characteristic of a particular register. Previous research had shown ambiguous findings regarding frequency. While some studies have found that use of high-frequency words decreases with age (Finn, 1977; Olinghouse &

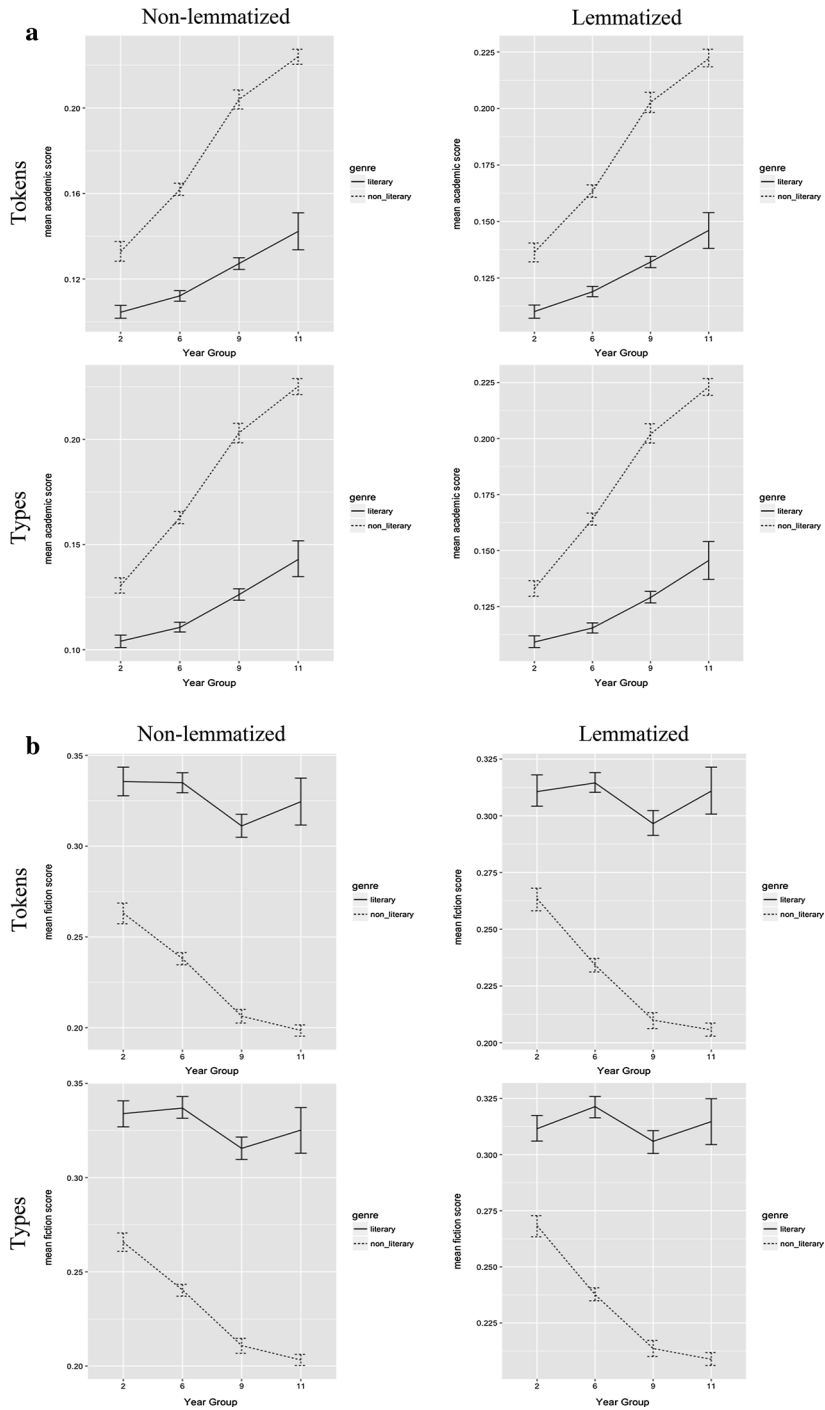


Fig. 4 Mean academic (a), fiction (b) vocabulary score

Table 16 Mixed-effects model for mean academic vocabulary score

	Value	SE	<i>df</i>	<i>t</i> -value	<i>p</i> value
(Intercept)	.1132	.0177	12.67	6.412	<.0001
Year group	.0049	.0016	181.40	2.965	.0034
Genre	.0230	.0184	237.00	2.120	.0350
Year group \times genre	.0033	.0014	307.80	2.319	.0210
Random effects	Variance			SD	
Writer within school			.0001		.0100
Title within discipline			.0007		.0263
School			.0006		.0248
Discipline			.0003		.0175
Residual			.0005		.0212
Goodness of fit					
R^2 marginal			.31		
R^2 conditional			.86		

Table 17 Mixed-effects model for mean fiction vocabulary score

	Value	SE	<i>df</i>	<i>t</i> -value	<i>p</i> value
(Intercept)	.3110	.0153	10.95	20.353	<.0001
Year group	-.0022	.0014	121.82	-1.518	.1316
Genre	-.0444	.0102	246.41	-4.354	<.0001
Year group \times genre	-.0039	.0014	296.65	-2.891	.0041
Random effects	Variance			SD	
Writer within school			.0001		.0080
Title within discipline			.0006		.0238
School			.0003		.0163
Discipline			.0003		.0163
Residual			.0006		.0248
Goodness of fit					
R^2 marginal			.41		
R^2 conditional			.80		

Graham, 2009; Olinghouse & Wilson, 2013; Sun et al., 2010), others either failed to find an effect (Malvern et al., 2004) or found that it applied only with certain groups (Lawton, 1963). The one study to look at overall mean frequency did not find differences between school children at different ages (Crossley et al., 2011). The present study also found that counts based on all lexical words did not show significant differences across year-groups or genres. However, in a more fine-grained analysis which separated the four lexical parts of speech, the mean frequencies of verbs and adjectives significantly decreased with age while the mean frequency of

nouns significantly increased. Importantly, frequency differences across age groups were significant only in analyses based on word tokens. When each distinct word in a text was counted only once, no such differences were found.

These findings, we have argued, imply that the standard model of vocabulary sophistication does not adequately capture vocabulary development in children's writing. When use in adult discourse is taken as the standard of frequency, younger children's writing differs from that of older children in that: (a) it frequently repeats nouns referring to entities that are rarely discussed in adult discourse; (b) it makes repetitive use of high-frequency verbs and adjectives. It is important to note that, while vocabulary sophistication has generally been seen as distinct from diversity, the fact that these developmental patterns cannot be expressed simply in terms of the repertoire of words used, but rather refer to extent of repetition implies that the former cannot be usefully separated from the latter. Lexical sophistication, in other words, should not be seen as an entirely distinct construct from lexical diversity.

Previous research has been mostly silent on the topic of register-appropriateness in children's vocabulary, with the most relevant strand of research being studies of 'academic' (Sun et al., 2010) or 'Greco-Latin' (Corson, 1985; Berman & Nir Sagiv, 2007) vocabulary, approaches which do not allow for a diversity of target genres in children's writing. The present study is therefore novel in attempting to model this aspect of sophistication. While our findings agree with previous research that use of typically 'academic' words increases as children mature, we also found that this increase was largely driven by non-literary writing. In literary writing, the increase was present, but more modest. Use of words typical of fiction texts (a category not studied by previous research), remained fairly constant across year groups while their use in non-literary texts decreased sharply.

It is not surprising that children's use of vocabulary becomes more register-appropriate as they progress through school. What is of more interest is that this development can be modelled in fairly simple quantitative terms and that such models appear to be a better index of development (as evidenced by the improved marginal R^2 s) than simple word-frequency-based measures. Analysis of vocabulary sophistication which do not take such register-related features into account appear, therefore, to be missing an important part of the developmental picture.

The central conclusion of this paper is that the relationship between vocabulary frequency and development in children's writing is far more complex than the simple equation of low-frequency with sophistication suggests. We have elaborated on this model by looking at how frequency interacts with part-of-speech, lexical diversity and register. It is unlikely that these elaborations exhaust the ways in which the model of vocabulary sophistication can be refined. Avenues which immediately suggest themselves for further exploration include integrating syntactic variables beyond simple part of speech analysis, and integrating phraseological analysis. Research in second language writing has shown categories such as collocation to be important aspects of development and to provide novel perspectives on learner language (e.g., Biber & Gray, 2013; Chen & Baker, 2016; Paquot, 2017). However, this work has been almost entirely ignored in studies of first language writing development. Collocation is important as it takes us beyond individual words to look at how words are used in relation to their co-text. It may be that much of the growth

in lexical sophistication lies in the relationships between the words which children use, rather than simply in what words they select. We noted above that the notion of appropriateness employed in this study is limited in that it addresses only the match between words and register, without considering correctness of use. Analysing the collocational contexts in which words are used may take us a step towards understanding appropriateness in this stronger sense.

Acknowledgements This research was supported by a Grant from the Economic and Social Research Council (ES/M00967X/1).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Aho, K., Derryberry, D., & Peterson, T. (2014). Model selection for ecologists: The worldviews of AIC and BIC. *Ecology*, 95, 631–636.
- Australian Curriculum and Assessment Reporting Authority. (2014). *F-10 curriculum: English*. Retrieved from <https://www.australiancurriculum.edu.au/f-10-curriculum/english/> 15th May 2018.
- Bartlett, E. J. (1984). Anaphoric reference in written narratives of good and poor elementary school writers. *Journal of Verbal Learning and Verbal Behavior*, 23, 540–552.
- Barton, K. (2018). *MuMIn: Multi-model inference, R package version 1.42.1*. Retrieved from <https://CRAN.R-project.org/package=MuMIn>. Accessed 17 Sept 2018.
- Berman, R. A., & Nir, B. (2010). The lexicon in writing-speech-differentiation. *Written Language and Literacy*, 13(2), 183–205.
- Berman, R. A., & Nir-Sagiv, B. (2007). Comparing narrative and expository text construction across adolescence: A developmental paradox. *Discourse Processes*, 43(2), 79–120.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D., & Gray, B. (2013). *Discourse characteristics of writing and speaking task types on the TOEFL iBT test: A lexico-grammatical analysis*. TOEFL iBT research report, 19.
- Carroll, J. B. (1964). *Language and thought*. Upper Saddle River, NJ: Prentice-Hall.
- Chen, Y. H., & Baker, P. (2016). Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Applied Linguistics*, 37(6), 849–880.
- Christie, F., & Derewianka, B. (2008). *School discourse: Learning to write across the years of schooling*. London: Continuum.
- Corson, D. (1985). *The lexical bar*. Oxford: Pergamon.
- Crossley, S. A., Roscoe, R., & McNamara, D. S. (2014). What is successful writing? An investigation into the multiple ways writers can write successful essays. *Written Communication*, 31(2), 184–214.
- Crossley, S. A., Weston, J. L., Sullivan, S. T. M., & McNamara, D. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication*, 28, 282–311.
- Davies, M. (2008–). *The Corpus of Contemporary American: 450 million words, 1990-present*. Retrieved from <http://corpus.byu.edu/coca/>. Accessed 1 May 2018.
- Davies, M. (2018). *Word frequency data*. Retrieved from <https://www.wordfrequency.info>. Accessed November 2012.
- Department for Education. (2014). *The National Curriculum in England. Framework document*. Retrieved from <https://www.gov.uk/government/publications/national-curriculum-in-england-framework-for-key-stages-1-to-4>. Accessed 22 April 2018.

- Department for Education. (2017). *Schools, pupils and their characteristics: January 2017*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/650547/SFR28_2017_Main_Text.pdf. Accessed 1 May 2018.
- Finn, P. J. (1977). Computer-aided description of mature word choices in writing. In C. R. Cooper & L. Odell (Eds.), *Evaluating writing: Describing, measuring, judging* (pp. 69–89). Urbana, IL: National Council of Teachers of English.
- Gardner, D. (2008). Validating the construct of word in applied corpus-based vocabulary research: a critical survey. *Applied Linguistics*, 28(2), 241–265.
- Garside, R., & Smith, N. (1997). A hybrid grammatical tagger: CLAWS4. In R. Garside, G. Leech, & A. McEnery (Eds.), *Corpus Annotation: Linguistic Information for Computer Text Corpora* (pp. 102–121). London: Longman.
- Gries, S. T. (2015). The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora*, 10(1), 95–125.
- Harley, J. (Ed.). (2018). *Why closing the word gap matters: Oxford language report*. Oxford: Oxford University Press.
- Koutsoftas, A. D., & Gray, S. (2012). Comparison of narrative and expository writing in students with and without language-learning disabilities. *Language, Speech and Hearing Services in Schools*, 43, 395–409.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26.
- Lawton, D. (1963). Social class differences in language development: A study of some samples of written work. *Language and Speech*, 6(3), 120–143.
- Malvern, D., Richards, B. J., Chipere, N., & Duran, P. (2004). *Lexical diversity and language development*. Basingstoke: Palgrave Macmillan.
- Massey, A. J., Elliott, G. L., & Johnson, N. K. (2005). *Variations in aspects of writing in 16+ English examinations between 1980 and 2004: Vocabulary, spelling, punctuation, sentence structure, non-standard English*. Cambridge: Cambridge Assessment.
- National Governors Association Center for Best Practices, C. O. C. S. S. O. (2010). *Common Core State Standards*. Retrieved from http://www.corestandards.org/wp-content/uploads/ELA_Standards1.pdf. Accessed 15th May 2018.
- Nesi, H., & Gardner, S. (2012). *Genres across the disciplines: Student writing in higher education*. Cambridge: Cambridge University Press.
- Olinghouse, N. G., & Graham, S. (2009). The relationship between the discourse knowledge and the writing performance of elementary-grade students. *Journal of Educational Psychology*, 101, 37–50.
- Olinghouse, N. G., & Leaird, J. T. (2009). The relationship between measures of vocabulary and narrative writing quality in second- and fourth-grade students. *Reading and Writing*, 22, 545–565.
- Olinghouse, N. G., & Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing: An Interdisciplinary Journal*, 26, 45–65.
- Paquot, M. (2017). The phraseological dimension in interlanguage complexity research. *Second Language Research*, 18, 1–25.
- Quigley, A. (2018). *Closing the vocabulary gap*. Abingdon: Routledge.
- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Rose, D., & Martin, J. R. (2012). *Learning to write/reading to learn: Genre, knowledge and pedagogy in the Sydney School*. Sheffield: Equinox.
- Sun, Y., Zhang, J., & Scardamalia, M. (2010). Knowledge building and vocabulary growth over two years, Grades 3 and 4. *Instructional Science*, 38, 147–171.
- Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics*. Boston, MA: Pearson.
- Uccelli, P., Dobbs, C. L., & Scott, J. (2013). Mastering academic language: Organization and stance in the persuasive writing of high school students. *Written Communication*, 30(1), 36–62.
- Zuur, A. F., Ieno, E. N., Saveliev, N. J., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. New York, NY: Springer.